The Southern Journal of Philosophy (1990) Vol. XXVIII, No. 3

THE MECHANICS OF RATIONALITY

Mark Leon
University of the Witwatersrand

I suspect philosophers fall into two categories, those with an overblown conception of our rationality, and those with an 'underblown' conception of it. Representatives of the latter variety, the deflaters, are the ones who set the agenda in the following. Though they do not succeed in burying our claims to rationality, they do succeed in forcing us to come up with a more realistic appraisal and a better understanding of the ways in which we do measure up against these claims.

The immediate challenge comes from Dennett (see especially Dennett 1984, though the challenge in the form of sphex goes back at least to Dennett 1973).[1] And the immediate general background to this challenge concerns the problem of reconciling our mechanicity and our rationality. There are at least two differentiable worries arising out of this. The one involves the thought that if at base our functioning is mechanical, then our rationality is undermined; rational and mechanistic systems are essentially in opposition to each other—where the one reigns the other is excluded. The other begins with mechanism and ends with the fear of our 'syntactitude'; the fear that we are syntactic engines the functioning of which requires no reference to *contentful* states of the kind considered essential to (our) rationality. Here, it is the first worry that I will be examining. The idea is to use it in order to put into perspective (an aspect of) the nature of our rationality and attendant limitations on it. As the discussion develops, the general attempt to exhibit the opposition between mechanism and rationality will be examined more directly.

Mark Leon is an Associate Professor of Philosophy at the University of the Witwatersrand. His research interests lie mainly in the philosophy of mind and epistemology. Articles of his have been published in the Australasian Journal of Philosophy, Mind and Language, Metaphilosophy *and* Philosophical Papers.

I.

First we set up the appearance of rationality; then we undermine it. Consider that contrivance that nature has thrown up to help us settle our thought experiments about rationality—sphex. Here I go through the main points in outline (for the details, see Wooldridge 1963, and Dennett 1984). The wasp sphex exemplifies a characteristic pattern when laying eggs. First it builds a burrow; then it seeks out and stings a cricket—enough to paralyse it but not to kill it; next it drags the cricket to the burrow and lays its eggs next to it. On hatching, the grubs are able to feed off the paralysed cricket. At first brush this is all very reasonable. It appears much less so when sphex is observed more closely. What sphex does is bring the paralysed cricket to the threshold of the burrow, leaving it there while 'checking' that all is well inside, and only then dragging it in. However, if the cricket is moved from its position on the threshold while the wasp is 'checking' within, it will run through the routine once more. And if the cricket is moved again sphex rehearses her performance again. If the routine is not performed endlessly, it is at least performed sufficiently to dissuade us from rating sphex's intelligence too highly. This feature of its behavior is clearly deep-wired, under strong stimulus control, inflexible and highly mechanical. Such behavior would tend to undermine rather than support the ascription of rationality.

So much for sphex: what about us? Viewed from one perspective our behavior can seem to be rational. But could it not be unmasked, viewed from a higher perspective? Admittedly, we are not exactly like sphex, but is there not a perspective from where we can be viewed which makes us seem far less rational and far more sphexish? For is our behavior not also a function of a mechanism albeit one of slightly greater complexity? Is our behavior not to some degree or other tropistic? Consider Dennett:

> 'The Godlike biologist reaches down and creates a slight dislocation in the wasp's world, revealing her essentially mindless mechanicity; could a superior intelligence, looking down on us, find a similar if more sophisticated trick that could unmask us?' (Dennett 1984, p. 11)

Sphex is illuminating; but mostly, I think, by contrast. Sphex's behavior is deep-wired. There is, one presumes, a mechanism which leads sphex to check that all is well within the burrow. What sets off the mechanism (presumably) is the bringing of the cricket to the threshold of the nest. What makes the pattern of behavior appear so unintelligent is the fact that the behavior is an automatism. It is not sensitive to other relevant factors or information. That the burrow has been checked makes no difference to the next occasion the 'checking response' is triggered.

We are not altogether unlike sphex. Firstly, our behavior too, I am willing to concede, is governed by a sort of mechanism. Secondly, whether that is true or not, we clearly have to admit that there is an inbuilt limit on the flexibility of our responses. These two points raise two corresponding questions. The general question is theoretically more important: if our behavior is at base governed by a mechanism, how can we be rational? How can we do what we do *because* reason dictates? The local point, if of less general interest, is nevertheless instructive—in asking how sphex-like we are, we are reminded of our shortcomings. Both these aspects will be examined in the following, as sphex is used to cast light on aspects of our rationality.

II.

Sphex constitutes the challenge. We can meet the challenge by sketching a necessary backdrop to the discussion. It is a common if still controversial contention that there are no *serious* laws of a psychological or (psycho-physical kind) under which mental events can be subsumed (see, for example, Davidson 1970 and 1973, and McGinn 1978 and 1979). The contention in its full generality is not the major interest here. I think there is something to the thesis, though not always that which others have sought or highlighted in it; but still it is limited in its scope. (And increasingly it comes to look as if the issue revolves on to how seriously we take 'serious'; whether we measure the psychological against the physical or more reasonably the biological; so often an aspect of our psychological functioning which is taken to fuel the thesis can be shown on reflection to be a feature shared with other non-basic sciences, like the biological (see Leon 1980)).

There is some reason to think that there could be, or should be, psychological laws. Psychological explanations share many important features with explanations normally taken to be law-like. For example, they support counter-factuals and subjunctive conditionals (see Leon 1980). When we explain an agent's behavior by reference to his reasons we are committed to holding that if he had failed to have those reasons he would not have acted that way; and if he had those reasons again, all other things being equal, he would act that way again.

However, even if we go along with this, how are we to take it? Is the suggestion that there is a law linking interesting couplets of reasons and actions? If that is so the account faces a serious, but instructive difficulty. There appears to be no hope of laws linking an agent's having reasons of a certain kind, with his acting in a specified way, let alone laws linking any agents having reasons of that kind with their acting in that specified way. Reason-action couplets do not generalise.

Why is this? Part of the problem here might just be that one is looking for laws at the wrong place; that one needs to look at the intentions one forms on the basis of one's reasons, rather than at reasons alone. The idea behind this is that while reasons might be presumptive considerations behind action, it is only once intentions are formed on the basis of them that we can take the agent to be sufficiently galvanised to act on them. Clearly we do need a distinction between having reasons for acting—reasons which predispose an agent to act—and reasons on which agents act—the reasons which determine an action. And this might be the right way or part of the right way of drawing it. But it hides the real problem.

The major factor which undermines the possibility of generalisations (the candidates for law-like status) linking couplets of reasons and actions concerns the way in which reasons issue in actions, namely, in the context of, or as mediated by, other reasons of the agent. While we might have reasons to perform certain actions, whether or not we perform them depends on other factors, like our other reasons. An agent might have reasons, say for eating, and yet fail to eat, for any number of reasons. An agent might have reasons for crossing the Rubicon but fail to cross it because of factors as diverse as 'structural ones'—those which philosophers tend to forget—like loss of courage to the possession of other, from the agents' point of view, 'better' reasons, for not acting that way or for doing something else. When we explain actions we do not assign actions to reasons taken one by one, we assign actions to reasons in the context of other reasons. How a person acts is a function of all, or a good many of, his beliefs and desires, their corresponding subjective probabilities and preference weightings, as well as a function of his values (if thought of as distinct from his desires).

Does this holism undermine the possiblity of any serious laws in this domain? It tells us this much: if you seek a law relating particular reasons to particular actions, it would have to be specified so as to apply strictly to an individual at a time. One might say: anyone with reasons R to do Ø would do Ø, *given* a set of background beliefs . . . and desires . . . and *given* a set of values . . . and *given* certain structural features, like a certain character . . . and *given* conditions of a certain kind . . . in effect we would be saying that an identical agent in identical conditions would, given the specified reasons act in the specified way. Is this a law-like statement? It is very high on specificity and very low on generality. The trouble is that when we explain an agent's actions by citing his reasons, we indicate that those reasons prevailed under conditions of competition. But though we would commit ourselves to holding that that agent under the same (internal and external) conditions would act in the same way again, we would not commit ourselves *unconditionally* to holding that that agent, let alone any agent, would act the same way, if he had the same reasons again. Agents are different and change, the environment changes . . . so long as reasons issue in actions only as mediated by other reasons, we cannot expect interesting couplets of reasons and actions. A reason enters as a possibility amongst other possibilities. Clearly certain reasons, perhaps those relating to the immediate well-being of agents have greater claims on one's attention, and so are more likely to affect one's intentions; but still the distinction remains between commanding one's attention and commanding one's assent.

If there is a law it would presumably refer to an agent's possession of reasons for acting and actions. Whereas an agent's having a particular reason to perform a particular action might speak little about what the agent would do next time possessed of that reason; his having a reason at least, under normal conditions, tells us that he will perform some action or other. There are grounds for expecting as it were a law of agency, perhaps of the following sort: whenever an agent A has reasons R to Ø, and no better reasons not to Ø, and no better reasons to do something else, and when he forms the intention to Ø for those reasons, and has the opportunity to Ø, he Ø-ies. This is a law by contrast with the previous offering, high on generality and low on specificity. It abstracts as it were from content. It links actioin to reason, not specific actions to content-specific reasons. It would not be very informative or predictive. But there is some chance that a generalisation of this form would be true.

Still there are instructive problems even with this suggestion. Consider the reference to an agent's 'best reasons.' This might be thought problematic in two ways. Firstly, as stressed by Davidson (Davidson 1973, p. 151) there is the problem of determining beforehand what reason is best, as opposed to determining after the event what reason has been operative. The schema does not, if this is correct, promise great empirical fruitfulness. There is something to this, but perhaps not quite what has been claimed. It is true that it is difficult to determine beforehand which reasons are an agent's strongest, especially so since there are perfectly legitimate ways of explaining away deviation from the expected. But for all that it does not follow that it is always impossible to determine what an agent's best reason is. Think of it this way. We associate certain sorts of reasons with certain persons. This is part of coming to know what a person is like. Yet we can be surprised by how people act, which is why, as in all cases of apparent falsification, there is work to be done in explaining (or explaining away) the deviation from the expected. We can revise our original estimate of what a person was like, and what reasons motivated him. Or we could discern some break or change from the past either as a function of new information or perhaps of maturation .... This sort of play-off indicates that though we cannot finally tie a person down, we have at least some sort of grasp prior to the performance of particular actions of what reasons motivate them, what reasons constitute their best reasons for acting. (Obviously, here we refer to reasons which require a more serious sort of judgment call.) Anyhow, it might be thought that this is a problem of an *epistemological* kind, involving our capacity to determine the initial conditions (the content and weightings of an agent's reasons); it is not clear why this epistemological consideration should affect the metaphysical status of the purported relation.

That leads to the second, and deeper point concerning the connected status respectively of 'best reason' and the so-called 'law of agency.' It does not seem to be an empirical (hence not falsifiable) truth, that agents act on, given the opportunity, their best all considered reasons. On the contrary, in normal conditions, the reasons an agent acts upon just are the reasons which, from the agent's point of view, are the best. (Here no judgment is passed on their universalisability, their intrinsic merit, or whether they are genuinely prudential for an agent.) We can handle deviations from such optimum functioning. But there is a limit on what can prevail without the subject losing his status as an agent.

In turn, there is a more general problem of which this is merely the symptom. Assume that we could come up with universal generalisations of sorts between reasons and actions. That still would not convince some that a law had been found. Not because such generalisations could not be sustained across possible worlds, not that is, because such generalisations would appear to be accidental. But on the contrary, because they would seem to embody a necessary truth, but one of the wrong, i.e., not *a posteriori*, character. Such generalisations have an *a priori* or conceptual character: reasons for actions just are states of an agent that would, all things being equal, issue in the corresponding actions (subject, of course, to modulation in the light of the agent's other states). Scientific laws do not have such a character. (The point is not that we could not discover that the law of agency failed of a particular candidate; rather the point is that any candidate of which it systematically failed would not be an agent. The empirical question is the range of application, not the nature of agency.)

There are many different and interesting aspects to the question concerning the nomological status of the relation between reasons and actions. It is not important to settle the issue here (though I touch on the issue once again in the conclusion to this paper). It is enough to have raised certain considerations which become important for the rest of my discussion. The most important feature to observe at this stage concerns the holistic way in which reasons issue in actions. We can agree on that—to some degree at least—even if we do not agree on the significance of this feature for the further question of the nomological status of reason explanations.

III.

So back to sphex. At first sight it might have been thought that sphex's behavior was reasonable. For reason does dictate that one should proceed cautiously and check one's path before acting—up to a point of course. But then sphex was so shamelessly unmasked and revealed to be acting not according to the dictates of reason but as a function of a deep-wired mechanism under tight stimulus control. Her behavior rather than being intelligent was shown to be a mere automatism. Or so it seems.

How sphex-like are we? It depends what is at issue. In the following I will argue that we are sphex-like in that like sphex, at base our behavior is governed by a mechanism—on one interpretation of that; still that we are mostly unsphex-like in that for the most part our behavior is not a product of such simple mechanism, and accordingly is not susceptible to the same degree of stimulus control; and that, thirdly, there are respects in which we are more interestingly sphex-like—differing from sphex only in degree.

Perspective comes by noting our differences from sphex. Afterwards the similarities will appear less threatening. So I begin with the second consideration, the respect in which we are significantly unsphex-like. I then move on to the third consideration before returning to the first.

The trouble with sphex is the simplicity of the controlling mechanism. Sphex's behaviour is governed by a mechanism under tight stimulus control, not subject to any feed-back in terms of previous findings, nor subject to moderation in the light of any other states enjoyed. Put sphex in the appropriate conditions and it is all too easy to trigger the response. Being deep-wired, inflexible and automatic, it is difficult on reflection to see this as an exemplification of intelligence.

Now recall the previous discussion concerning the connection between reasons and actions. The crucial point was that reasons, beliefs and desires, issued in actions only in the context of, and as moderated by other reasons, other beliefs and desires of an agent. What is exhibited is that the serious ascription of belief and desire begins where stimulus bound action, as in, say, reflexes, ends. (It is notable that with sphex there is no separating out, even conjecturally, what sphex might reasonably be said to desire and what sphex might reasonably be said to believe.) Where stimulus bound action tapers off—where there is no systematic mapping from initial states to ensuing behaviour—holism comes into play. Given an initial state, say a desire, whether is gets acted upon depends as well on the other concurrent states of the agent. This is the pattern of the normal course; our initial states *set the agenda* when a course of action is to be determined, but they do not alone *determine the outcome.*

Hence the variations even given the same agenda-setting states. Certain sorts of characters, or characters at different stages of their lives, when having reasons to cross the(ir) Rubicon might or might not cross it. More interestingly, (in this context) whether they cross it depends on what other beliefs might be held, what other desires are enjoyed (or suffered), what other values are held. So even if we could associate with a reason, a specific and determinate course of action, we could not generalise from possession of the reason to performance of the action.

This is not to suggest, taken to the extreme on the other side, that there is a pure egalitarianism of initial states or that external occurrences comprise input of a similarly non-differentiable quality. Clearly, even in the case of normal functioning certain internal states, certain desires or needs of an agent command more attention than others, for example, those states which affect the survival or immediate well-being of an agent. And similarly, certain (external) stimuli are more likely to produce a 'corresponding' effect than others. Guns are reasonably good instruments for not only capturing attention, but also quite typically capturing assent. Or combining these aspects, the smell of a favoured dish especially to the recently deprived, can elicit an almost sphex-like response. (And here I do not refer to salivation.) Nevertheless, 'commanding one's attention,' or 'being higher up on the agenda' do not add up to 'compelling certain actions.' The more outcome is determined in interaction with other factors, with other desires and beliefs, and the more valuation is brought into play, the more rational the behavior is.

IV.

There is, accordingly, a connection between what fails in the case of reason-action relations and our relative freedom from sphexishness. But to make this point we do not need to represent ourselves as being more exclusively rational than we are; nor as being free from mechanism altogether. By their nature mechanisms enable certain things to occur, while limiting the range of outcomes. In sphex's case the mechanism is very simple. In our case, if our operation were governed by one, it would be of greater sophistication and complexity. But it would still be in its own way limiting.

We need to pursue this point further. Think of our normal limitations. How we act is a function of what alternatives we are able to represent to or for ourselves. But what we are in this way able to represent is limited. Our repertoire is all too finite. How we act is also a function of what inferences we might draw from our initial beliefs, and what principles of inference we might use. But in these regards too we are respectively limited and less than optimally 'rational.'[2] We are not all knowing. And our reasoning capacity is not ideal. And though ideally our beliefs ought to be a function solely of evidence, clearly there are factors other than evidential ones which affect what we believe. On the conative or motivational side, we are also less than rationally ideal. How we act is a function of the strength or importance of some of our initial states, or our needs and desires, and values. Some might dominate when 'reason' would have suggested otherwise. Even if our cognitive powers were not limited our performance would depart from a rational ideal.

But these respects in which we fall short of some idealised rationality do not alter the fact or suggest other than that we remain properly even if not fully rational. Rationality just is in part the operation of a system like ours; the determination of action by reasons, beliefs and desires (with the associated reasoning operations) in interaction with the other internal states of the agent. We rate sphex less on the rationality scale because her responses are not moderated by her other states; *not* because her responses are governed by a mechanism.

It is true that holistic operation constitutes in some sense an ideal. (Perhaps an ideal at a certain level of functioning. There is reason to think that the mind is to some extent modular; certainly at the level of perceptual input systems (see Fodor 1983). Altogether more interesting would be to find modules (as alluded to below) at the level of central processing—the cognitive facility that receives (and possibly reciprocally affects to a certain extent) the deliverances of the input systems). In practice we wander to some degree or other away from the holistic ideal. In the abnormal case we begin to cross over into more specifically sphex-like territory and practice. But even before abnormality is reached, a certain amount of fall-off from the ideal can be tolerated. If the ideal of rationality, even relative to our resources, consists in all relevant factors coming into play in determining action (or in the acquisition of belief) we fall short of that ideal. Mostly less than all relevant factors come into play and there need be nothing pathological about that. To a certain extent the relevant factors are compartmentalised. A certain amount of compartmentalisation is not only tolerated but presumably required. For the end of thought is often action; and endless reflection in the light of all information would preclude action. Compartmentalisation is also a limited solution to the problem of the lack of complete unity and integration of an agent's character and projects. There is, however, a question as to how much compartmentalisation can be tolerated while operating with the notion of a single, coherent, agent. In the normal condition lies the seeds of a certain sort of abnormality.

A similar fall off from the ideal of full holistic accounting occurs with habitual action as opposed to reasoned deliberative action. But again it is easy to see how this might be justified. For agents to survive under conditions in which quick action can be demanded, for agents to act rather than be prone to inertia, for there not to be cognitive overload, a certain economy in deliberation is called for. And found.

Not all departures from the ideal are like the foregoing normal (in moderation). Consider, for example, the senile or the brain-damaged. In these cases actions might be under the control of only parts of the intentional system. In the extreme case we begin to wonder if there is even half a person operating; whether there is any intentional control at all.

A different fall off from holistic operation involves the inequality of initial states. There is reason to believe (and indeed reason to expect) certain needs and desires are to some degree deep-wired and (normally) dominant. All things being equal their writ runs. Such states would at least feature high on the agenda, and would most closely approximate states with a nomological force. Reasons to eat tend to end in eating. But even states of this kind, with increased weight because of their importance to the survival of the agent—hence the (evolutionary) reason for their being deep-wired—do not compel action. There are some who would die rather than eat, and have died rather than eat. Again the more the outcome is determined in part or balance by other relevant factors, the more rational the behavior is. Even where there is an initial unequal weighting of states.

But as an extension of this normal inequality there is a pathological condition. Consider for example compulsions and obsessions. These are internal states possibly triggered by external occurrences or stimuli which are more tightly connected to behavior. With compulsions the internal states no longer (merely) set the agenda; they in effect directly determine action. It is not necessarily that there is no consultation with other relevant factors; the point is that such states can override their (normally) moderating influences. Viewed this way compulsions are like automatisms, actions of a more or less automatic, unreasoned and inflexible kind. The degree to which they dominate rather than being moderated by the agent's other states (or considerations) constitutes a degree to which they are beyond the agent's control. (One can generalise: the more deliberation in the light of all relevant factors, the more reasoned an action, the more voluntary or free the action is. Freedom here is akin to (as a necessary condition at least) stimulus autonomy. The less deliberation in the light of other factors, the less moderation by other factors, the more automatic and inflexible the response is, the less voluntary or free it is. What is at issue is not total freedom from control; but the type of control. Though my immediate concern is with rationality rather than freedom, one gets a glimpse of how the two are connected. That is as it should be. There is a strong line of thought suggesting that our greater freedom (in contrast to other animals and indeed children) is a function of our greater deliberation or at least capacity for deliberation (where necessary). The cutoff is not between, as Descartes might have thought, mechanism and its absence. The cutoff is between types of mechanisms and what their differential flexibility and moderation respectively permit. (On the connection between freedom and reason, see Zemach 1987, p. 270.))

There is of course much else that distinguishes sphex from us; there is much else that counsels against the serious ascription of beliefs and desires to sphex. Crucially beliefs get ascribed where stimulus control tapers off; and stimulus control tapers off where stimuli constitute input into a system of internal states (of a subject). It is precisely where a smooth mechanical correlation between physical stimuli and 'behavioral' responses falls off, that one seeks enlightenment in 'meaningful' or 'contentful' states—ways in which the environment is represented to, or for, the subject. And once we do that the typical holistic moderation comes into play, and with it our ability to separate out beliefs and desires. This is to scratch the surface.

No matter. The issue here was: how threatening is mechanism to rationality? The answer given is that it all depends on the nature of the mechanism. I have yet to concede (formally) that what sphex does is not rational at all. I think it is (with much needed qualifications to come concerning what exactly is being maintained here). But the crucial point is that if we treat sphex as one pole in a spectrum of cases, we can plot the respects in which we are like or unlike sphex. In doing this we plot the respects in which we are more or less rational; for rationality is a matter of aspect and degree. It is the (relative) flexibility of our system, and the way in which our states are moderated by our other states which demarcates us from sphex. It is the complexity and sophistication of the mechanism which is crucial. So even if, as with sphex, our behavior can be unmasked, in the sense that we can be shown to be only loosely or imperfectly fitted to our environment, with a significantly limited repertoire of responses, it would still remain true that we are rational—to a degree. (Indeed that we are rational to that degree we might previously have thought. For sphex really throws up nothing new. The example simply allows us to see certain of our features in a better perspective.)

V.

So much for the local skirmish with the mechanist theory. Still to be faced head-on is the general worry that mechanism threatens rationality. If mechanism is true how could it also be true that we do what we do because of reason; that we do what we do because reason dictates?

In answering this question, once again it is worth starting with sphex, with something less close to home. But what is the worry? Consider Dennett again:

> 'Sphex . . . typically does what reason dictates, for surely reason does proclaim the wisdom of reconnoitering her pathways . . . But she doesn't do it because reason dictates it *to her*; it is rather that her designed behaviour cunningly approximates in this one regard the responsiveness of what we might call . . . Pure Practical Reason.' (Dennett 1984, pp. 25-26).

The 'text' is suggestive more than embodying a specific and determinate statement of the problem. But what it suggests covers what I take to be part of the central problem. For the moment disregard the first part. The suggestion in the second is that at best you can get a mechanism, by means of say, an external co-ordination solution, to approximate a system acting according to the dictates of reason, without really acting according to the dictates of reason. It is designed to imitate what reason would dictate; but lacking reason it does not do what it does because reason so dictates (to it). Fleshing this out further, the idea is that the principles governing a mechanism are not, being intrinsic, the principles governing a rational system. After all, the result is produced by the mechanical operation of the system. At that level no reference is made to the "dictates of reason." So at best a mechanism can imitate but not reproduce or duplicate the operation of reason.

The thesis that the operations of reason and mechanism are mutually exclusive is badly mistaken. But before demonstrating that we need to concede the obvious. Does sphex do what she does because reason dictates? Clearly not—in one sense at least. Sphex lacks explicit reasons for doing what she does. She has no reason and so no reason to dictate to her. However, and this is where sphex is illuminating, there is *a* reason (or purpose) for her doing what she does, not merely a cause. If sphex lacks reason it does not follow that her actions lack reason (on an important interpretation of that). On the contrary, the truth is that there is a reason for what she does, and moreover a mechanism which ensures that that writ of reason runs.

These points can be made clearer by distinguishing a weaker and stronger sense in which something can be done because reason dictates. In the *weak* sense a system acts because of the dictates of reason if there is at least a coordination mechanism which ensures that what reason dictates gets done or is acted upon. On this weak reading the system need not have its own reasons, and so the explanation for what is done makes no reference to the system's autonomous set of reasons. It is enough that there is a clear sequence and connection between what 'reason decrees' and what is done. A system can act because of the dictates of reason without acting for its own reasons. In the *strong* sense a system does what it does because of the dictates of reason only if the reasons it acts upon are its own. If something is done for a reason, or if, say, a belief is acquired as a consequence of reasoning, then the action or belief occur because the system's own reason dictates. The writ of reason which runs is its own.

Sphex unlike us has no reason of her own.[3] If she embodies (a low-level) rational system she does so only in the weak sense. Sphex has a mechanism which enables her to do what reason dictates, to act at the behest of a reason. It is in sphex's interest, conducive to her or her progeny's survival, to check her path before proceeding. What reason would dictate for a system with reason of its own, nature ensures by means of a mechanism. In sphex's case the mechanism operates coarsely, with obvious attendant deficiencies and limitations.

There are those who would, probably, more readily grant that a system like sphex could do what reason (on some interpretation) would dictate, *when* reason so dictated, while not granting that the system did it *because* reason dictated. The problem it would seem consists in the jump from 'doing what is done *when* reason dictates' to 'doing what is done *because* reason dictates.' How can reason be effective?

The answer, in the first place, is to put aside the question of the autonomy or internalisation of reason. That will be dealt with later. The problem after that has a philosophically, even if not technically, easy solution. One moves from a *when* to a *because* by putting the response onto a systematic footing. The shortfall in merely doing something *when* reason dictates is that the connection is or could be accidental or coincidental. For reason to dictate is for reason to be operative; for what was done not merely to coincide with what reason dictates. Hence the solution; the systematic connection between what reason dictates and what is done.

How credible is this? Should the solution require that the system have a voice of reason of its own? I think not. Not in the first place where one's interest is in getting the system to do what reason does dictate. There are different ways in which the 'writ of reason' can run. Some, for example strong rational systems, will be more sophisticated than others. But I cannot see how one can deny that the writ of reason does run if there is a system which systematically does what reason dictates. The systematic connection ensures that the covariance is not coincidental. It is not coincidence that what the system does is what reason dictates. But is that not to exhibit that that reason dictates the outcome enters into the very production of that outcome? The system operates that way *because* reason dictates. How the mechanism accomplishes what it does is not the issue. What is important is what the mechanism operating (normally) *enables*—in this instance the execution of what reason dictates. (If it turned out that sphex went into her routine because some kindly but not very bright soul pushed a button in the specified conditions, that would not undermine the claim that what was done was done because reason dictated. Rather that would be a highly extravagant solution to a fairly simple (as these things go) design or engineering problem. The kindly soul would be—to the extent that he was making reason's voice incarnate—part of the system's mechanism.)

Consider a parallel but this time intentionally contrived case, a simple calculator. When one successively pushes the *2* button, the *multiplication* button, and the *5* button, *10* is flashed on the screen. Why? The answer: because 5 × 2 = 10. That 5 × 2 = 10 is an explanation of why *10* flashed on the screen. For the point of the device is to give the appropriate answer to the selected or essayed mathematical problem. It is not an accident it gives the answer it does; for that was part of its designed function.

Of course to say that *10* flashed because that was the answer to the problem is not to give a complete or exhaustive answer. There is still the mechanical reply. When those buttons are pushed in that sequence, that pattern of lights results. But does this reply undermine or make irrelevant the 'rational' response? Surely not. The sequence of button pushings would not have had the observed lighting results if the designer of the system had not had the multiplication problem (amongst others) in mind. That, after all, was the rationale for the system's operation. But that rationale is physically realised in the way the system is hard-wired. The two explanations go hand in hand. The hard-wiring mechanical operation *enables* the 'rational' operation. Without having its own reasons (and so without being itself intelligent), insofar as the device (when operating normally) systematically gives the right answer, it makes (a fragment of mathematical) reason incarnate. There is a reason for its flashing *10*, and that reason is the fact that 5 × 2 = 10. To account for its behaviour we need to make reference to that fact. To focus solely on the mechanical side of the operation is to miss what that mechanical operation enables—*enables, not* undermines.

To summarise: if it is systematically true that a system or device does b when b is what reason dictates—it is not a matter of coincidence—then it is true that the device does b because reason dictates. There is a coordination solution which ensures that b is done when reason decrees. Where there is a systematic connection between what reason decrees and what is done, that reason (so) decrees enters into the explanation of what is done. But that just is for the system to do b *because* that is what reason dictates. Satisfying reason is not an accidental feature of it; it is built or engineered to satisfy reason.

Neither sphex nor calculators have reason. They are systems which instantiate weak rationality. They do what reason dictates without possessing reason—mindlessly as it were. We, on the other hand, do have reason. How can our possession of reason be compatible with our mechanicity? How can the operation of our reason require—as I will be arguing—mechanism?

Consider the lead from systems of weak rationality. Such systems can be seen as engineering solutions, natural or contrived, to coordination problems—coordinating action with reason. With systems of weak rationality, the voice of reason is external; nature's (of course only so to speak) or the intentional contriver's. With systems of strong rationality the voice of reason lies within. Such reason is autonomous or underived. With systems of strong rationality there are two central ways in which reason plays a role; as internal states of the system, beliefs and desires, producing behaviour, and as internal states of the system, typically belief, producing other internal states, for example, other beliefs.

With systems of weak rationality the voice of reason is external; the system is so 'engineered'—has a certain intrinsic functioning—that it does what this external reason dictates. The efficacy of reason is built into the system by shaping the intrinsic mechanism in accordance with, an imagined or real, task designation. With systems of strong rationality, the voice of reason is internal; the systems have certain reasons for doing something or certain reasons for believing something. How do we get this internal writ of reason to run? At the very least when we act or believe for a reason there is a sequence of states; pre-existing internal states, and the subsequent action or newly generated (or further maintained) internal states. At most that sequence alone tells us that the system does or believes b when reason dictates. How do we get from 'doing b *when* reason dictates' to 'doing b *because* reason dictates'? The answer has clearly already been foreshadowed. (And once again one must not confuse its apparent philosophical simplicity with a corresponding technical simplicity. In all likelihood implementing such a system is beyond us.) The answer is to provide a systematic connection between the internal reason and the resulting event. And that systematic connection, as before, can be a function of any appropriate sort of mechanism.

We need to take this slowly. The first point is this: if our actions or some of our internal states were products of a mechanism which mediated (we see with hindsight) the 'wishes of the voice of reason' in some systematic way, that would be a perfectly good implementation of the writ of reason. The subsequent states would occur because of, hence be explained by, the prior reasons. Mechanism does not (in itself) undermine the writ of reason; mechanism of an appropriate kind *enables* the writ of reason to run.

The second point concerns possible limitations on the 'appropriateness' of the connection. I would not be inclined to put much in the way of constraints on this notion. It would have to be the case (in normal conditions and relative to 'design' limitations) that when we have reason to do, or believe something, we do or believe as prescribed. When sequences of that sort are systematically implemented, then I would take it to be the case that (part of) the explanation for that sequence is the fact that reason so dictates. The sequence would not be accidental or coincidental. Accordingly I would count as suitable a mechanism which took in another agent, so long as the other agent was constrained, for whatever reason, by the requirement that the writ of the original agent runs. An interesting consequence of this might be that there is a limit on the degree to which an agent might be fooled into thinking that he really is the agent of his own actions—when he is not. The better he 'is fooled' the more evident it becomes that the perpetrator is himself the victim or handmaiden of the agent. After all, he is enabling or ensuring that the writ of that original agent runs.

The third point concerns the nature of the underlying mechanism. Clearly there is a massive jump when we move from systems which lack (their own) reason, to systems which possess such reason. Coincident with this jump is a movement away from behaviour or operation under tight and fairly automatic stimulus control, to behaviour or operation exhibiting the noted features of holistic moderation. Clearly these differences would be reflected in the nature and complexity of any underlying enabling mechanism. My objective (and obligation) here is not to supply information about how the relevant intrinsic mechanisms would function; in the first place it is rather to do two things, exhibit that if there were such underlying mechanisms which 'mediated' reason in either sense, such mechanisms would not thereby negate the voice of reason, and secondly (to exhibit) that there is a natural continuum between the operation of sphex-like creatures and us, differences turn not on the presence or absence of mechanism, but the nature of the mechanisms and what they respectively enable.

It might be thought that so far, at best, I have shown that a mechanism would suffice, not that it is necessary. What would exhibit its necessity? Though I think there is no alternative it is worth pursuing this a bit further here. (More of a general character is added in the concluding section). Consider the question: When does a system have reason? Does it have reason when, believing p and where q follows from p, it comes to believe q? Or does it have reason only when, believing p and believing q follows from p, it comes to believe q? In general does a system have reason when it has a sequence of beliefs which parallel sequences of well accepted inferences? Or must it also explicitly believe the principle of inference? Those who argue against the mechanistic account might go for the latter on the grounds that a reasoning process is not a mere, even systematic, sequence of states, but a sequence of states underpinned by a belief in the principle of inference.

The trouble is two points are being mixed up here: the first concerns the relative sophistication of the device; the second concerns the question of 'mechanism.' As to the first point, reason or its possession cannot be uniquely specified by either alternative. The difference is one of level of sophistication. (There is an interesting side question: Does a system operating according to the first schema have, if not an explicit belief, at least a tacit or implicit belief? But this will not be taken up here.) As to the second point it is evidently a mistake to think that if there is no inference when a belief of one kind systematically follows a belief of another kind, that there is such inference simply by adding a higher order belief in the principle of inference (unless, more reasonably, the point is to exhibit the bedrock for reason). Any rational process requires implementation in intrinsic material—hardware. Even where reason, principles of inference, become explicit—believing q because believing q follows from p—a mechanism is still needed which ensures the transition. Adding a higher order belief is not an alternative solution (to getting the inference to run); it simply pushes the problem back one point further. Implementation mechanisms are structures which carry subjects from one state to the next, when the latter follows from the former. They are mechanical solutions to coordination problems. Without them there might be logical relations but no real rational process realised. More generally: rule-governed operations at some point have to be implemented. It cannot be that each instance of a rule is subsumed under a higher-order rule. In that direction lies a regress; not a real operation. (See Carroll 1895.)

Reference to the implementation of a reasoning system is a good point at which to bring together the threads of the discussion. As calculators and computers exhibit, perhaps too well, the implementation of reasoning relations, is realised by a system with intrinsic states, the sequence of which is designed to mirror the corresponding inferential relations. The point might be generalised: rational systems are systems whose mechanical operations by mirroring the 'relations of reason' enable the writ of reason to run.

Rational systems are engineering solutions to design problems. The efficacy of reason turns on whether (or the degree to which) the design objectives are realised. Indulge in a little fantasy. Assume you want to construct a system with a certain pattern or operation; certain internal states making appropriate other internal states as well as behaviours—a system with beliefs and desires productive of other beliefs and also behaviour. The task then is to build the 'efficacy of reason' into the system by shaping an intrinsic mechanism in accordance with the designated, blue-print, requirement. The system has two levels of operation—of description and explanation. The trick is to coordinate the two levels. Conceptually prior is what we want the system to do; we want a system such that when one state 'makes reasonable another' that other occurs. It is part of the blueprint for the system that it exhibits or instantiates the rational pattern. The second step is to come up with a mechanism which enables the realisation of the rational pattern. Its operation would ensure that when the system is in the prior state, the state which makes appropriate the other, that second state follows.

Initial qualms aside, a central problem is this. Have we not constructed a system the causal efficacy of which lies with the mechanism and not with its apparent reasons? How can the reasons be genuinely explanatory as opposed to merely epi-phenomenal? Why does the writ of the supervenient base not usurp all authority?

This is no place to deal with all the challenges posed by supervenience of higher-order functioning on lower-order mechanical functioning, but a sketch of a response is required and can be given. The efficacy of reason may be thought of as safe for two reasons (in the case of strong rational systems). Firstly, the device is constructed so that its intrinsic functioning in paralleling the rational relations thereby enables those relations to occur. Even if the mechanism is viewed as the motor, the motor serves the higher-order relations or objectives. (Think again of the calculator; the sequence of intrinsic states is explained by (draws its rationale from) the fact that that sequence encodes, parallels, the mathematical relations.) Similarly, now generalised, the sequence of intrinsic states (the supervenient base of the rational systems) draws its rationale from what it mirrors thereby from what it enables. The very functioning of the mechanism is determined by the higher-order pattern we want realised. And as suggested in the earlier reference to paths of operation going through other agents, it is difficult to see how the successful operation of the mechanism can at the same time be a failure—the failure to secure the efficacy of reason—for successful operation systematically connects what reason ordains with what occurs. Failure comes rather when there is a breakdown in the normal functioning of the mechanism.

With strong rational systems, the mechanism not only draws the rationale from the fact that it enables the writ of reason to run, more specifically its function is to enable the writ of the internal voice of reason to run. With a strong rational system what has to be coordinated is the system's internal states, its reasons and its subsequent actions. A strong rational system is rational not only in form (the implementation of a rational sequence—doing b *because* directed by reason) but also in content—it has its own reasons. But once again to the extent that the mechanism functions successfully, how could it fail to give substance to this added dimension; ensuring the efficacy of the system's internal states?

Fantasy has its place. But it must not overtake reality. Of course we are not engineered and certainly not engineered so as to realise a pre-defined pattern of functioning. But that is to get the point the wrong way round (in our case). We happen to be so structured as to realise (to some approximation) a rational pattern. We are not the products of nature's deliberate design; but we could have been the products of deliberate design. We are, however this came about, an engineering solution to a design problem. What an engineer might contrive, nature in her blind way can throw up (assuming, as one need not, aspects of evolutionary theory). Whether or not we are the products of nature's deliberation, nature (within limits) smiles kindly on us because the way we behave confers on us a selective advantage. Similarly, if in some distant and isolated part of the universe matter were to come together in such a way as to comprise what we could employ as a calculator, that confluence of matter would not have any intended function at all. Nevertheless it would be such as to enable, when pressed into service, a certain outcome, namely, calculations. Whether that was its function that is what it would enable. What is crucial is not the objective behind the device but the objective fact of what it does or would enable. Rational functioning can 'emerge' from intrinsic functioning.

VI.

Concluding comments on rationality and mechanisms:

In some ways it is a little too easy to diffuse the 'fears of our sphexishness' even when granting certain similarities to sphex. And this route might hide significant differences arising from the added sophistication of our rational functioning. To redress this possible distortion I conclude with a few comments to provide a more general perspective.

Putting aside the threat from sphex it might be thought that I have done less than justice to the worry about mechanism. Consider, for example, the following: With rational systems (at least with strong ones), where rational explanation is called for, there is an internal, not merely external, connection between the subject's reasons and the event, belief or behaviour, explained; for the event to be explained is explained in virtue of the fact that it is made reasonable in the light of the explanatory factor, the subject's reason. In short, the event to be explained and the explanatory factor are linked together non-intrinsically. But then how can a mechanical account—a set of intrinsic connections—capture the notion of 'rational appropriateness'?

This 'objection' is well made. And I think the point has to be taken. However to take the point is not to adopt an anti-mechanist position. Indeed, there is no alternative mechanism (understood as conflicting with an intrinsic mechanism) that can be extracted from the thought. We need to accommodate the insight without jettisoning the only way of making reason effective. And the insight can be accommodated by recognising that rational explanation is a species of teleological explanation. It shares with teleology more generally the non-intrinsic relational connection between the event to be explained and the event explaining it. With teleology an event is explained in virtue of the fact that it is (possibly put too strongly) *required* for a certain end. It occurs because it is so required. This is no mere juxtaposition of intrinsically characterised events. Similarly, I would argue with rational explanation an event is explained in virtue of the fact that it is (possibly put too weakly) *appropriate* in the light of a certain reason. But in both cases—I think this is most clearly evident with teleology[4]—this non-intrinsic relation (at the one level) can be (indeed I would argue, must be) underpinned by an intrinsic mechanism at a different level. Once again more generally, intrinsic mechanisms are the means to realising non-intrinsic relations.

This feature of the operation of reason connects up with a point made earlier. The thesis that there is an internal connection between beliefs and what they explain when they rationalise, that is, make reasonable, goes hand in hand with the thesis that the nature of the reason-action relation is known *a priori* not *a posteriori*. Beliefs are events (which in coordination with desires) make reasonable certain other events. Being a belief and being such as to rationalise are not accidental features. (This is not to suppose that beliefs do not cause what they rationalise; it is rather to suggest that there is more than (only) causation connecting events when one is rationalised by another.) If the consequence is that rationalisation is not a species of ordinary law-like explanation, we at least now see or glimpse how and why that is; as well as seeing what other company (ie., teleological explanation in general) such explanation keeps.

One last point to conclude. There is a third thesis (which I have maintained all along) which goes hand in hand with the preceding two; suitably understood there is no opposition between rational and mechanical explanation. Of course, any purported rational explanation can be falsified, by discovering a simpler explanation which better accords with the facts. Hence the obvious reluctance to accord sphex too high a rationality. There is not much in the way of compelling reason to ascribe beliefs and desires to sphex. There are however, those who think any mechanistic theory undercuts a rational one. That is wrong. It might be that for any purported rational explanation there is an apparently equally good mechanistic explanation of the same event. But we need to get clear on what that involves. There are those mechanistic explanations which as indicated falsify the rational; and there are those which provide an explanation of the event to be explained without obviously falsifying the rational. They provide an explanation in different terms. Such explanations do not falsify. They do not even make the rational redundant (at best a case of 'over determination'); what they do is underpin. That is the point at which rationality and mechanism connect.

The more sophisticated the rationality of the system, the more sophisticated the mechanism. Suitably understood (or operating) there is no opposition between rational and mechanical systems. They are doing different jobs. And in terms of the jobs done need to be appropriately matched.[5]

NOTES

[1] It is not Dennett's position, but the underlying question, that I am primarily interested in. Still, it is worth noting Dennett's 'apparent' position. He buys, I think, the premises of the argument but resists or opts for a different conclusion. The premises suggest that we are always off-centre because the mechanism can at most approximate (or imitate) rationality; but whereas that might lead some to believe that we are not rational, Dennett instead seems to take it as given that we are rational, and so that is what our rationality comes to. Second best is good enough. My position is that what the argument (or the mechanist's scenario) shows is that we are incompletely or imperfectly rational; but not that we are not properly rational.

[2] Concerning the limitations on our reasoning capacities, see the highly interesting even if somewhat controversial work of Tversky and Kahneman (for example, Tversky and Kahneman 1974).

[3] At least it does no harm to my argument to interpret sphex in this minimal way.

[4] Teleological functioning exhibits the same distinction as rational functioning with respect to weak and strong systems. Weak systems are goal-directed, but serve the goals of others (so to speak). Strong systems are goal directed, serving their own goals. The *form* of explanation is the same—both appeal to teleological laws, doing b because it is "required" for g. The difference is that with strong systems *goals* enter into the *content* of the explanation.

[5] I am grateful to Michael Pendlebury for some helpful comments. I also acknowledge the financial assistance of the Institute for Research Development of the Human Sciences Research Council towards this research; but of course the opinions expressed in this paper and the conclusions drawn are my own.

REFERENCES

Carroll, L., 1895. 'What the tortoise said to Achilles,' *Mind*, Vol. 4.

Davidson, D., 1970. 'Mental Events' in *Experience and Theory*, eds. L. Foster and J. W. Swanson, London, Duckworth.

Davidson, D., 1973. 'Freedom to Act' in *Essays on Freedom of Action*, ed. T. Honderich, London, Routledge and Kegan Paul.

Dennett, D. G., 1973. 'Mechanism and Responsibility' in *Essays on Freedom of Action*, ed. T. Honderich, London, Routledge and Kegan Paul.

Dennett, D. G., 1984. *Elbow Room: The Varieties of Free Will Worth Wanting*, Cambridge, MA. The MIT Press/A Bradford Book.

Fodor, J., 1983. *The Modularity of Mind*, Cambridge, MA. The MIT Press.

Leon, M., 1980. 'Are Mental Events Outlaws?', *Philosophical Papers*, Vol. IX.

McGinn, C., 1978. 'Mental States, Natural Kinds and Psychophysical Laws,' *Proceedings of the Aristotelian Society* Supp. Vol. LII.

McGinn, C., 1979. 'Action and its Explanation' in *Philosophical Problems in Psychology*, ed. N. Bolton, London, Methuen.

Tversky, A. and Kahneman, D., 1974. 'Judgment under Uncertainty: Heuristics and Biases,' *Science*, Vol. 185.

Woolridge, D., 1963. *The Machinery of the Brain*, New York, McGraw Hill.

Zemach, E., 1987. 'The Makings of Mind,' *The Southern Journal of Philosophy*, Vol. XXV.